

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**

4 日本国特許庁

PATENT OFFICE  
JAPANESE GOVERNMENT

02.06.00

JP00/3623

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日  
Date of Application:

1999年 7月 7日

27 JUL 2000

出願番号  
Application Number:

平成11年特許願第193141号

出願人  
Applicant(s):

セイコーエプソン株式会社

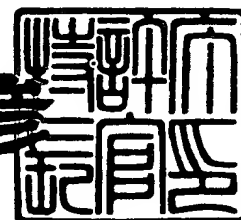
PRIORITY  
DOCUMENT

SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH RULE 17.1(a) OR (b)

2000年 6月29日

特許庁長官  
Commissioner,  
Patent Office

近藤隆彦



出証番号 出証特2000-3052055

【書類名】 特許願

【整理番号】 J0074156

【提出日】 平成11年 7月 7日

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 7/24  
G06F 17/30

【発明の名称】 情報分類方法及び情報分類装置並びに情報分類処理プログラムを記録した記録媒体

【請求項の数】 24

【発明者】  
【住所又は居所】 長野県諏訪市大和3丁目3番5号 セイコーエプソン株式会社内  
【氏名】 長石 道博

【特許出願人】  
【識別番号】 000002369  
【氏名又は名称】 セイコーエプソン株式会社  
【代表者】 安川 英昭

【代理人】  
【識別番号】 100093388  
【弁理士】  
【氏名又は名称】 鈴木 喜三郎  
【連絡先】 0 2 6 6 - 5 2 - 3 1 3 9

【選任した代理人】  
【識別番号】 100095728  
【弁理士】  
【氏名又は名称】 上柳 雅誉

【選任した代理人】  
【識別番号】 100107261  
【弁理士】

【氏名又は名称】 須澤 修

【手数料の表示】

【予納台帳番号】 013044

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9711684

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 情報分類方法及び情報分類装置並びに情報分類処理プログラム  
を記録した記録媒体

【特許請求の範囲】

【請求項 1】 複数の文書に対しそれぞれの文書の共通性に基づいてクラスタリング処理し、それによって得られたクラスタリング結果に対し、クラスタリング結果の概要を表すクラスタリング結果概要一覧表を作成し、そのクラスタリング結果概要一覧表を前記クラスタリング結果とともに出力するようにしたことを特徴とする情報分類方法。

【請求項 2】 前記クラスタリング結果概要一覧表は、クラスタリング処理されることによって得られたそれぞれのクラスタのクラスタ名によって構成されることを特徴とする請求項 1 記載の情報分類方法。

【請求項 3】 前記クラスタリング結果とクラスタリング結果概要一覧表とを相互にリンクし、クラスタリング結果概要一覧表の所定のクラスタ名部分が指示されると、前記クラスタリング結果の対応するクラスタ部分の表示を可能とし、クラスタリング結果の或るクラスタ部分を指示すると、クラスタリング結果概要一覧表の表示を可能とすることを特徴とする請求項 2 記載の情報分類方法。

【請求項 4】 前記クラスタリング結果概要一覧表の或るクラスタ名部分が指示されることによって、前記クラスタリング結果内の対応するクラスタ部分の表示を行う際、そのクラスタを囲む枠の最上部またはそれより 1 つ手前に存在するクラスタの枠内に存在する最終行部分を先頭にした表示を行うことを特徴とする請求項 3 記載の情報分類方法。

【請求項 5】 前記クラスタリング結果の或るクラスタ部分が指示されることによって、前記クラスタリング結果概要一覧表の表示を行う際、クラスタリング結果概要一覧表の先頭部分からの表示を行うことを特徴とする請求項 3 記載の情報分類方法。

【請求項 6】 前記クラスタリング結果概要一覧表を構成するそれぞれのクラスタ名の配列順序は前記クラスタリング結果内のクラスタ配列順序に従うことを特徴とする請求項 2 から 5 のいずれか 1 項に記載の情報分類方法。

【請求項 7】 前記クラスタリング結果概要一覧表の表示を行う際、前記クラスタリング結果のそれぞれ対応するクラスタの重要度に応じて、クラスタリング結果概要一覧表上におけるクラスタ名の表示の仕方を変えることを特徴とする請求項 2 から 6 のいずれか 1 項に記載の情報分類方法。

【請求項 8】 クラスタリング対象となる複数の文書が、ユーザの与えたキーワードによって検索された文書である場合、ユーザの与えたキーワードを含むクラスタ名は前記クラスタリング結果概要一覧表上においてそのクラスタ名の表示の仕方を他のクラスタ名とは異ならせることを特徴とする請求項 2 から 7 のいずれか 1 項に記載の情報分類方法。

【請求項 9】 複数の文書に対し、それぞれの文書の共通性に基づいてクラスタリング処理するクラスタリングモジュールと、

このクラスタリングモジュールによって得られたクラスタリング結果に対し、クラスタリング結果の概要を表すクラスタリング結果概要一覧表を作成する概要一覧表作成部と、

そのクラスタリング結果概要一覧表を前記クラスタリング結果とともに出力する表示制御モジュールと、

を有することを特徴とする情報分類装置。

【請求項 10】 前記クラスタリング結果概要一覧表は、クラスタリング処理されることによって得られたそれぞれのクラスタのクラスタ名によって構成されることを特徴とする請求項 9 記載の情報分類装置。

【請求項 11】 前記クラスタリング結果とクラスタリング結果概要一覧表とを相互にリンクし、クラスタリング結果概要一覧表の所定のクラスタ名部分が指示されると、前記クラスタリング結果の対応するクラスタ部分の表示を可能とし、クラスタリング結果の或るクラスタ部分を指示すると、クラスタリング結果概要一覧表の表示を可能とすることを特徴とする請求項 10 記載の情報分類装置。

【請求項 12】 前記クラスタリング結果概要一覧表の或るクラスタ名部分が指示されることによって、前記クラスタリング結果内の対応するクラスタ部分の表示を行う際、そのクラスタを囲む枠の最上部またはそれより 1 つ手前に存在

するクラスタの枠内に存在する最終行部分を先頭にした表示を行うことを特徴とする請求項 11 記載の情報分類装置。

【請求項 13】 前記クラスタリング結果の或るクラスタ部分が指示されることによって、前記クラスタリング結果概要一覧表の表示を行う際、クラスタリング結果概要一覧表の先頭部分からの表示を行うことを特徴とする請求項 11 記載の情報分類装置。

【請求項 14】 前記クラスタリング結果概要一覧表を構成するそれぞれのクラスタ名の配列順序は前記クラスタリング結果内のクラスタ配列順序に従うことを特徴とする請求項 10 から 13 のいずれか 1 項に記載の情報分類装置。

【請求項 15】 前記クラスタリング結果概要一覧表の表示を行う際、前記クラスタリング結果のそれぞれ対応するクラスタの重要度に応じて、クラスタリング結果概要一覧表上におけるクラスタ名の表示の仕方を変えることを特徴とする請求項 10 から 14 のいずれか 1 項に記載の情報分類装置。

【請求項 16】 クラスタリング対象となる複数の文書が、ユーザの与えたキーワードによって検索された文書である場合、ユーザの与えたキーワードを含むクラスタ名は前記クラスタリング結果概要一覧表上においてそのクラスタ名の表示の仕方を他のクラスタ名とは異ならせることを特徴とする請求項 10 から 15 のいずれか 1 項に記載の情報分類装置。

【請求項 17】 複数の文書に対しそれぞれの文書の共通性に基づいて文書の分類を行って出力する情報分類処理プログラムを記録した記録媒体であって、その情報分類処理プログラムは、

複数の文書に対しそれぞれの文書の共通性に基づいてクラスタリング処理する手順と、

それによって得られたクラスタリング結果に対し、クラスタリング結果の概要を表すクラスタリング結果概要一覧表を作成する手順と、

そのクラスタリング結果概要一覧表を前記クラスタリング結果とともに出力する手順と、

を含むことを特徴とする情報分類処理プログラムを記録した記録媒体。



【請求項 18】 前記クラスタリング結果概要一覧表は、クラスタリング処理されることによって得られたそれぞれのクラスタのクラスタ名によって構成されることを特徴とする請求項 17 記載の情報分類処理プログラムを記録した記録媒体。

【請求項 19】 前記クラスタリング結果とクラスタリング結果概要一覧表とを相互にリンクし、クラスタリング結果概要一覧表の所定のクラスタ名部分が指示されると、前記クラスタリング結果の対応するクラスタ部分の表示を可能とし、クラスタリング結果の或るクラスタ部分を指示すると、クラスタリング結果概要一覧表の表示を可能とすることを特徴とする請求項 18 記載の情報分類処理プログラムを記録した記録媒体。

【請求項 20】 前記クラスタリング結果概要一覧表の或るクラスタ名部分が指示されることによって、前記クラスタリング結果内の対応するクラスタ部分の表示を行う際、そのクラスタを囲む枠の最上部またはそれより 1 つ手前に存在するクラスタの枠内に存在する最終行部分を先頭にした表示を行うことを特徴とする請求項 19 記載の情報分類処理プログラムを記録した記録媒体。

【請求項 21】 前記クラスタリング結果の或るクラスタ部分が指示されることによって、前記クラスタリング結果概要一覧表の表示を行う際、クラスタリング結果概要一覧表の先頭部分からの表示を行うことを特徴とする請求項 19 記載の情報分類処理プログラムを記録した記録媒体。

【請求項 22】 前記クラスタリング結果概要一覧表を構成するそれぞれのクラスタ名の配列順序は前記クラスタリング結果内のクラスタ配列順序に従うことを特徴とする請求項 18 から 21 のいずれか 1 項に記載の情報分類処理プログラムを記録した記録媒体。

【請求項 23】 前記クラスタリング結果概要一覧表の表示を行う際、前記クラスタリング結果のそれぞれ対応するクラスタの重要度に応じて、クラスタリング結果概要一覧表上におけるクラスタ名の表示の仕方を変えることを特徴とする請求項 18 から 22 のいずれか 1 項に記載の情報分類処理プログラムを記録した記録媒体。

【請求項 24】 クラスタリング対象となる複数の文書が、ユーザの与えたキーワードによって検索された文書である場合、ユーザの与えたキーワードを含むクラスタ名は前記クラスタリング結果概要一覧表上においてそのクラスタ名の表示の仕方を他のクラスタ名とは異ならせることを特徴とする請求項 18 から 23 のいずれか 1 項に記載の情報分類処理プログラムを記録した記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は汎用の検索サービスで検索された結果に対しクラスタリング処理を施し、それによって得られたクラスタリング結果に対しクラスタリング結果の概要を表す一覧表を作成して、クラスタリング結果とともにその一覧表の表示をも可能とした情報分類方法及び情報分類装置並びに情報分類処理プログラムを記録した記録媒体に関する。

【0002】

【従来の技術】

ネットワーク上に存在する膨大な量の情報の中からユーザの所望とする情報を検索する場合、検索サービスの存在は重要である。たとえば、インターネットで web ページを検索する際、ユーザは、幾つかの検索サービスの中から任意の検索サービスを選び、自分の欲しい情報を得るための検索要求としてのキーワードを入力する。これによって、検索サービス側では、入力されたキーワードに基づいて情報検索を行って、その検索結果をユーザに提示する。

【0003】

しかし、検索サービスによって検索される情報は膨大な量となることも多く、その中からユーザの本当に欲しい情報を見つけるのは非常に大変である。近年、web ページは増大の一途をたどっているため、検索された多数の情報を如何にユーザにわかりやすく提示するかが大きな課題となっている。

【0004】

最近では、検索された情報をユーザが見やすい形に加工して提示する手法も開発され実用化されつつある。たとえば、ユーザの入力したキーワードで検索され

た結果から得られるキーワードを用いて再検索することで、情報の絞り込みを行い、ユーザの所望とするwebページを見つけやすくする方法がある。つまり、検索によって得られる検索結果の集合を特徴づけるキーワードを抽出して、ユーザの本当に欲しい情報の集合に収束させる方法である。

#### 【0005】

このように、膨大な情報の中から、似た情報の集合を見つけることをクラスタリングという。情報処理ではこのクラスタリングはよく知られた手法であり、膨大な文書を分類する場合などに広く使われている。

#### 【0006】

##### 【発明が解決しようとする課題】

クラスタリングされて得られた結果（クラスタリング結果という）は、一般には、表形式で提示される。図2にクラスタリング結果の一例を示す。このクラスタリング結果は、ユーザが「半導体」というキーワードをある検索サービスに与えたとき、その検索サービスから得られた多数の文書に対しクラスタリング処理した結果である。

#### 【0007】

この図2で示されるクラスタリング結果は、クラスタリングによって得られたそれぞれのクラスタ（共通性を有する文書群の1つの集合）の名称（クラスタ名という）と、それぞれのクラスタに属する検索結果、それぞれのクラスタに幾つの文書が含まれているかを示す文書数などが一覧表形式で示されている。このようにクラスタリング結果は表形式で提示されるのが普通である。この図2の場合、クラスタリング処理によって、「概要」というクラスタ名を持つクラスタ、「LP (Laser Printer)」というクラスタ名を持つクラスタ、「仕様」というクラスタ名を持つクラスタ、「デバイス」というクラスタ名を持つクラスタ、「半導体」というクラスタ名を持つクラスタ、「電子」というクラスタ名を持つクラスタの6個のクラスタが得られることになり、それぞれのクラスタには、共通性を有する複数の文書が存在する。

#### 【0008】

この図2に示されるクラスタリング結果の例では、クラスタリングによって得

られたクラスタの数は6個とそれほど多くはなく、ユーザはそのクラスタリング結果全体を把握するのにそれ程多くの時間を費やさずに済む。しかし、クラスタリングによって得られたクラスタの数は時として何十個あるいは何百個といった膨大な数量となる場合もある。このような場合、クラスタリング結果全体を見るだけでも大変である。したがって、自分の欲しい情報がどのクラスタに有りそうかを調べるには、画面上に表示されたクラスタリング結果を上から順につぶさに見て行くという作業が必要であり、多くの時間と労力が費やされるという問題がある。

#### 【0009】

そこで本発明は、汎用の検索サービスで得られた検索結果をクラスタリング処理を施し、かつ、クラスタリングによって得られたクラスタリング結果全体の概要を一目で把握できるような一覧表を作成することで、ユーザが自分の欲しい情報を効率よく探すことができるようにすることを目的としている。

#### 【0010】

##### 【課題を解決するための手段】

前述の目的を達成するために、本発明の情報分類方法は、複数の文書に対しそれぞれの文書の共通性に基づいてクラスタリング処理し、それによって得られたクラスタリング結果に対し、クラスタリング結果の概要を表すクラスタリング結果概要一覧表を作成し、そのクラスタリング結果概要一覧表を前記クラスタリング結果とともに出力するようにしている。

#### 【0011】

また、本発明の情報分類装置は、複数の文書に対し、それぞれの文書の共通性に基づいてクラスタリング処理するクラスタリングモジュールと、このクラスタリングモジュールによって得られたクラスタリング結果に対し、クラスタリング結果の概要を表すクラスタリング結果概要一覧表を作成する概要一覧表作成部と、そのクラスタリング結果概要一覧表を前記クラスタリング結果とともに出力する表示制御モジュールとを有する構成としている。

#### 【0012】

また、本発明の情報分類処理プログラムを記録した記録媒体は、複数の文書に

対しそれぞれの文書の共通性に基づいて文書の分類を行って出力する情報分類処理プログラムを記録した記録媒体であって、その情報分類処理プログラムは、複数の文書に対しそれぞれの文書の共通性に基づいてクラスタリング処理する手順と、それによって得られたクラスタリング結果に対し、クラスタリング結果の概要を表すクラスタリング結果概要一覧表を作成する手順と、そのクラスタリング結果概要一覧表を前記クラスタリング結果とともに出力する手順とを含むものである。

【0013】

これら各発明において、前記クラスタリング結果概要一覧表は、クラスタリング処理されることによって得られたそれぞれのクラスターのクラスター名によって構成されている。

【0014】

そして、前記クラスタリング結果とクラスタリング結果概要一覧表とを相互にリンクし、クラスタリング結果概要一覧表の所定のクラスター名部分が指示されると、前記クラスタリング結果の対応するクラスター部分の表示を可能とし、クラスタリング結果の或るクラスター部分を指示すると、クラスタリング結果概要一覧表の表示を可能としている。

【0015】

そして、クラスタリング結果概要一覧表の或るクラスター名部分が指示されることによって、前記クラスタリング結果内の対応するクラスター部分の表示を行う際、そのクラスターを囲む枠の最上部またはそれより1つ手前に存在するクラスターの枠内に存在する最終行部分を先頭にした表示を行うようにしている。

【0016】

また、クラスタリング結果の或るクラスター部分が指示されることによって、前記クラスタリング結果概要一覧表の表示を行う際は、クラスタリング結果概要一覧表の先頭部分からの表示を行うようにしている。

【0017】

また、前記クラスタリング結果概要一覧表を構成するそれぞれのクラスター名の配列順序は、前記クラスタリング結果内のクラスター配列順序に従うようにしてい

る。

【0018】

さらに、クラスタリング結果概要一覧表の表示を行う際、前記クラスタリング結果のそれぞれ対応するクラスタの重要度に応じて、クラスタリング結果概要一覧表上におけるクラスタ名の表示の仕方を変えるようにしている。

【0019】

さらに、クラスタリング対象となる複数の文書が、ユーザの与えたキーワードによって検索された文書である場合、ユーザの与えたキーワードを含むクラスタ名は前記クラスタリング結果概要一覧表上においてそのクラスタ名の表示の仕方を他のクラスタ名とは異ならせるようにしている。

【0020】

このように本発明は、膨大な情報をクラスタリング処理することによって得られたクラスタリング結果の概要が一目でわかるようなクラスタリング結果概要一覧表をクラスタリング結果とともに表示するようにしたので、クラスタリング結果が膨大な内容となったとしても、クラスタリング結果全体のおおよその内容を把握することができ、ユーザが自分の欲しい情報を探す際の大きな助けとなる。これによって、ユーザは自分の欲しい情報を効率よく探すことができ、また、クラスタリング結果全体のおおよその内容を把握できることから、思わぬ情報の存在を発見する可能性もある。

【0021】

そして、クラスタリング結果概要一覧表は、クラスタリング処理されることによって得られたそれぞれのクラスタのクラスタ名によって構成される。すなわち、クラスタリング処理を行う方法の1つとして、たとえば、複数の文書のそれぞれのタイトルにどのような特徴的要素があるかを調べ、文書のタイトルに共通する特徴的要素を含む文書を同じクラスタとするという方法があるが、このようにしてクラスタリング処理されて得られたクラスタリング結果には、それぞれクラスタごとにクラスタ名が付される。したがって、このクラスタ名を取り出して、それを用いてクラスタ結果概要一覧表とすることが可能であり、それによって作成されたクラスタリング結果概要一覧表は、いわばクラスタリング結果の見出し

的なものとなり、その一覧表を構成するクラスタ名を見れば、クラスタリング結果にはどのような内容の情報が存在するかのおおよその見当が付き、自分の欲しい情報を探すのが容易となる。

#### 【0022】

また、クラスタリング結果とクラスタリング結果概要一覧表とを相互にリンクすることで、クラスタリング結果概要一覧表からクラスタリング結果の対応するクラスタ部分に簡単に移行することができ、また、その状態から、クラスタリング結果概要一覧表に容易に戻ることができるので、クラスタリング結果が膨大な内容であっても、このような操作を何回か行うことで、やがては欲しい情報の入っているようなクラスタを探すことができ、自分の欲しい情報を効率よく探すことができる。

#### 【0023】

そして、クラスタリング結果概要一覧表の所定のクラスタ名部分を指示したときクラスタリング結果の対応するクラスタ部分の表示を行う際、対応するクラスタを囲む枠の最上部またはそれより1つ手前に存在するクラスタの枠内に存在する最終行部分を先頭にした表示を行うようにしているので、本来表示すべきクラスタのクラスタ要素（文書タイトルなど）が確実に表示がなされるようになる。

#### 【0024】

同様に、クラスタリング結果の所定のクラスタ部分を指示したときクラスタリング結果概要一覧表を表示する際、クラスタリング結果概要一覧表の先頭部分からの表示を行うようにしているので、一覧表の内容が多い場合や一覧表が複数作成される場合であっても、ユーザが画面上でクラスタリング結果を見ていて、その画面をクラスタリング結果一覧表に戻した場合、一覧表全体を見渡すことができるため、戻った位置が一覧表全体のどの位置かがわからなくなることがなくなる。

#### 【0025】

また、クラスタリング結果概要一覧表を構成するそれぞれのクラスタ名の配列順序は、前記クラスタリング結果内のクラスタ配列順序に従うようにしている。つまり、クラスタリング結果内のクラスタが、たとえば、それぞれのクラスタ内

に含まれる文書数の多い順の配列となっていれば、その順序でクラスタリング結果概要一覧表のクラスタ名を配列する。

【0026】

一般に、クラスタリング結果のクラスタ配列順序は何らかの意味を持っているので、クラスタリング結果概要一覧表のクラスタ名の順序もそれに従うようにした方がユーザが情報を探す上で都合がよい場合が多い。たとえば、クラスタリング結果におけるクラスタ配列順序がユーザの与えたキーワードに対する適応度の高い順となっているとすれば、クラスタリング結果概要一覧表におけるクラスタ名の配列順序もその順序に従って配列する。ユーザは一般に上から順に見て行く傾向にあるので、そのような配列とすれば欲しい情報を効率よく探すことができる。

【0027】

また、前記クラスタリング結果概要一覧表の表示を行う際、前記クラスタリング結果のそれぞれ対応するクラスタの重要度に応じて、クラスタリング結果概要一覧表内のクラスタ名の表示の仕方を変えるようにしている。ここでいうクラスタの重要度というのは、たとえば、ユーザの与えたキーワードに対する適応度の高さなどであり、キーワードに対する適応度が高いクラスタに対応するクラスタ名部分は他のクラスタ名とは異なった色での表示を行ったり、その大きさを変えて表示したり、ブリンク表示とするなどの措置を講ずる。これによって、そのクラスタリング結果概要一覧を見るだけで、自分の欲しい情報がどこに入っているかのおおよその見当を付けやすくなる。

【0028】

さらに、ユーザの与えた検索キーワードを含むクラスタ名の表示の仕方を他のクラスタ名とは異ならせることによっても、自分の欲しい情報がどこに入っているかのおおよその見当を付けやすくすることができる。特に、自分の与えたキーワードそのものを含むクラスタ名を有するクラスタには自分の欲しい情報の入っている可能性が高いので、このような機能を設けることは使い勝手の向上に大きく寄与できるものとなる。



【0029】

## 【発明の実施の形態】

以下、本発明の実施の形態について説明する。なお、この実施の形態で説明する内容は、本発明の情報分類方法および情報分類装置についての説明であるとともに、本発明の情報分類処理プログラムを記録した記録媒体における情報分類処理プログラムの具体的な処理内容をも含むものである。

【0030】

図1は本発明を実現するための装置構成を示すもので、大きく分けると、検索サービス1、変換モジュール2、クラスタリングモジュール3、クラスタリング結果概要一覧表作成モジュール（以下、概要一覧表作成モジュールという）4、表示制御モジュール5とから構成され、変換モジュール2、クラスタリングモジュール3、概要一覧表作成モジュール4、表示制御モジュール5が本発明の情報分類装置に相当し、特に、概要一覧表作成モジュール4を設けたことに本発明の特徴がある。

【0031】

検索サービス1はインターネットなどで一般に広く使用されている汎用の検索サービスであり、ユーザからの検索要求としてのキーワードが入力されることにより、たとえばwebページなどから、入力されたキーワードに基づいた情報検索を行う。この検索サービス1で検索された検索結果はファイル形式で出力され、クラスタリングモジュール3に渡されるが、複数の検索サービスが存在する場合、それぞれの検索サービスによって出力されるデータ形式が異なるので、どのような検索サービスからのファイルであってもその内容を読めるような形式にファイルを変換するために変換モジュール2が設けられる。

【0032】

クラスタリングモジュール3は、検索サービス1により出力された検索結果（変換モジュール2による変換後のファイル内容）に対しクラスタリング処理を行うもので、この実施の形態では、それぞれの文書から文書のタイトルを抽出し、そのタイトルに含まれる特徴的な用語を特徴要素として抽出し、抽出された特徴要素に基づいてクラスタリング処理を行う。

## 【0033】

具体的には、それぞれの文書からタイトルとして抽出された部分を形態素解析し、形態素解析された結果から特徴的な用語を特徴要素として抽出する。その後、抽出された特徴要素とそれぞれの文書との関係を示す特徴テーブルを作成する。この特徴テーブルは、たとえば、抽出されたそれぞれの特徴要素が各文書のタイトルに幾つ含まれているかを、それぞれの特徴要素とそれぞれの文書と対応づけて示すもので、一例として、それぞれの文書のタイトルから、「概要」、「LP」、「仕様」、「デバイス」、「半導体」、「電子」というような特徴要素が抽出されたとすると、これらの特徴要素が、それぞれの文書のそれぞれのタイトルに、それぞれ何個含まれているかを示す内容となっている。

## 【0034】

このような特徴テーブルに基づいて、多数の文書を意味的に共通性のある複数のクラスタに分類する。つまり、それぞれの文書のそれぞれのタイトルに存在する特徴要素に基づいて、タイトルの中に共通する特徴要素を持つ文書を1つのまとまりとし、そのまとまりを1つのクラスタとする。

## 【0035】

このクラスタリングモジュール3から、たとえば、図2のようなクラスタリング結果が出力されたとする。この図2は前述したように、クラスタリングされることによって得られた各クラスタの名称（上述の特徴要素に相当し、それをここではクラスタ名と呼んでいる）と、そのクラスタに属するそれぞれの文書のタイトルと、そのクラスタには幾つの文書が含まれているかを示す文書数、さらには、それぞれのタイトルごとにスコアを示す数値などが表形式で示されている。

## 【0036】

なお、このスコアは、与えられたキーワードとそれぞれの文書との適合度を示す客観的な尺度として用いられ、ここでは、値が大きいほど、与えられたキーワードに適合する内容を有する文書であるとする。

## 【0037】

そして、このクラスタリングモジュール3によって得られたクラスタリング結果におけるそれぞれのクラスタの配列は、図2に示されるように、ここでは、そ

れぞれのクラスタに含まれる文書の数の多い順となっている。前述したように、この場合、上から順に、概要クラスタ、LPクラスタ、仕様クラスタ、デバイスクラスタ、半導体クラスタ、電子クラスタの順となっている。

#### 【0038】

概要一覧表作成モジュール4は、クラスタリングモジュール3から出力されたクラスタリング結果に基づいて、それぞれのクラスタのクラスタ名に基づいて、クラスタリング結果の概要を表すクラスタリング結果概要一覧表（概要一覧表という）を作成するものである。

#### 【0039】

表示制御モジュール5は、クラスタリングモジュール3によって得られたクラスタリング結果や、概要一覧表作成モジュール4で作成された概要一覧表をユーザに提示するために様々な表示制御を行うもので、この実施の形態では、クラスタリング結果とともに概要一覧表を表示させることは勿論、クラスタリング結果と概要一覧表のリンク部分の相互の表示制御、さらには、注目すべきクラスタをユーザの目に付きやすい表示とするといった様々な表示制御を行う。その具体的な表示制御内容については後述する。

#### 【0040】

このような構成において、本発明の情報分類処理について説明する。本発明が行う情報分類処理手順は概略的には、図3のフローチャートに示すように、まず、汎用の検索サービス1で検索された検索結果を取得し（ステップS1）、取得した検索結果に対しクラスタリング処理を施し（ステップS2）、そのクラスタリング結果を出力する（ステップS3）。そして、そのクラスタリング結果に基づいて概要一覧表作成処理を行い（ステップS4）、作成された概要一覧表を上述のクラスタリング結果とともに表示する（ステップS5）。なお、この作成された概要一覧表を上述のクラスタリング結果とともに表示する際の表示の仕方としては、たとえば、クラスタリング結果に概要一覧表を画面上で重ね表示するようにしてもよく、また、概要一覧表とクラスタリング結果を別個に配置し、概要一覧表の後に続いてクラスタリング結果を表示させるようにしてもよい。なお、クラスタリング結果の容量が多い場合には、画面をスクロールすることによっ

て、クラスタリング結果を順次画面上に現すようにする。

#### 【0041】

以下、本発明の情報分類処理内容について具体例を参照しながら詳細に説明する。

#### 【0042】

この実施の形態では、クラスタリングモジュール3が行うクラスタリング処理は、検索サービス1で検索された文書に対し、それぞれの文書のタイトルを抽出し、そのタイトルから特徴要素を抽出して、抽出された特徴要素とそれぞれの文書との関係を示す特徴テーブルを作成して、その特徴テーブルの内容に基づいて、それぞれの文書を意味的に共通性のある複数のクラスタに分類する。また、この実施の形態では、ユーザが「半導体」というキーワードを検索要求として検索サービス1に入力し、それによって得られた多数の文書がクラスタリングモジュール3によってクラスタリングされ、そのクラスタリング結果が図2に示すような結果であったとする。

#### 【0043】

このクラスタリングモジュール3からのクラスタリング結果は、概要一覧表作成モジュール4に入力され、以下に示すような処理がなされる。

#### 【0044】

まず、図2で示されるクラスタリング結果における各クラスタ（概要クラスタ、LPクラスタ、仕様クラスタ、デバイスクラスタ、半導体クラスタ、電子クラスタ）において、それぞれのクラスタ名（「概要」、「LP」、「仕様」、「デバイス」、「半導体」、「電子」）を用いた概要一覧表を作成し、その概要一覧表をクラスタリング結果とともに表示する。図4は概要一覧表10をクラスタリング結果20とともに表示した例を示すもので、この表示例では、概要一覧表10に続いてクラスタリング結果20を表示した例である。ここで用いたクラスタリング結果20は、クラスタ数が6個ときわめて少ない例であるが、実際には、クラスタ数は何十個あるいは何百個というような場合もあり、その中から自分の欲しい情報を探すには、クラスタリング結果全体を見ながら探すことになるので、単に、クラスタリング結果全体を見せられただけでは自分の欲しい情報を探す

のに大変な労力を要する。これに対し、クラスタ名による概要一覧表が表示されることによって、その概要一覧表を構成するクラスタ名を見渡すことで、クラスタリング結果にはどのようなクラスタがあるのか、自分の欲しい情報はその中のどのクラスタに入っていそうかといったことをおおよそ見当付けることができる。

#### 【0045】

そして、概要一覧表10を構成するそれぞれのクラスタ名とクラスタリング結果とはそれぞれ対応する部分がリンクされている。したがって、たとえば、クラスタリング結果のクラスタ数が多く、一つの画面上ですべてのクラスタリング結果が表示しきれないような場合でも、図4で示される概要一覧表10の任意のクラスタ名をクリックすると、それに対応するクラスタリング結果20のクラスタ部分が即座に表示され、その状態で、今度はクラスタリング結果のクラスタ名をクリックすると概要一覧表に即座に戻るといった表示制御が可能となる。

#### 【0046】

このような表示制御を行う際、使い易さをより一層向上させるために次のような機能を付加する。

#### 【0047】

まず、概要一覧表におけるクラスタ名の配列順序は、クラスタリング結果におけるそれぞれのクラスタの配列順序に従う。すなわち、図2のクラスタリング結果を例に取れば、そのクラスタリング結果におけるそれぞれのクラスタの配列順序は、それぞれのクラスタに含まれる文書数の多い順となっている。たとえば、概要クラスタには文書数は16個、LPクラスタにも文書数は16個、仕様クラスタには文書数は14個、デバイスクラスタには文書数は9個、半導体クラスタには文書数は7個、電子クラスタには文書数は4個といった具合である。したがって、この場合、概要一覧表においても、図4に示すように、一覧表の左から右に、「概要」、「LP」、「仕様」、「デバイス」、「半導体」、「電子」の順とした配列とする。

#### 【0048】

なお、クラスタリング結果におけるクラスタ順位の付け方としては、文書数だ

けではなく、それぞれのクラスタのスコアに基づいて決めることもできる。図2のクラスタリング結果を見ると、前述したように、そのクラスタに属するそれぞれの文書のタイトルと、そのクラスタには幾つの文書が含まれているかを示す文書数、さらには、それぞれのタイトルごとにスコアを示す数値などが表形式で示されている。

#### 【0049】

このスコアは、与えられたキーワードに対しその検索サービスが独自の方法で検索したとき、検索された文書対応に付けられた値であり、一般には、与えられたキーワードとそれぞれの文書との適合度を示す客観的な尺度として用いられ、汎用の検索サービスによる検索結果にはこのスコアが付されるのが普通である。このスコアは、検索方式により計算の仕方や値の考え方などが異なるものの、一般に、値が大きいほど、与えられたキーワードに適合する内容を有する文書であるといえる。

#### 【0050】

したがって、それぞれのクラスタごとにそれぞれのクラスタにおけるスコアの平均などを求め、その平均のスコアの大きいクラスタ順に配列することも考えられる。このように、クラスタリング結果におけるそれぞれのクラスタ順位を、平均のスコアの大きいクラスタ順とした場合には、それによって作成される概要一覧表のクラスタ名の順序もそれに従う。

#### 【0051】

たとえば、概要クラスタの平均のスコア（このスコアはキーワードとの適合度を表すものであるので、その単位としては、%や点数など検索サービスによって異なるがこの実施の形態では点数で表すものとする）が約36点と求められ、LPクラスタの平均のスコアが約28点と求められ、仕様クラスタの平均のスコアが29点と求められ、デバイスクラスタの平均のスコアが約94点と求められ、半導体クラスタの平均のスコアが108点と求められ、電子クラスタの平均のスコアが87点と求められたとする。このように計算された平均のスコアを各クラスタのスコア（クラスタスコアと呼ぶ）とする。そして、このクラスタスコアの高い順にクラスタの順位を再構成すると、この場合、クラスタスコアの最も高いク

ラストは、半導体クラスタの108点であり、第2位はデバイスクラスタの94点であり、第3位は電子クラスタの87点であり、以下、概要クラスタ（36点）、仕様クラスタ（29点）、LPクラスタ（28点）といった順序となる。

#### 【0052】

このように、クラスタリング結果におけるそれぞれのクラスタの配列順序が、それぞれのクラスタのクラスタスコアの高い順となっているとすれば、それによって作成される概要一覧表のクラスタ名の順序も、クラスタリング結果のクラスタ配列順序に従って、「半導体」、「デバイス」、「電子」、「概要」、「仕様」、「LP」の順序とする。

#### 【0053】

一般に、クラスタリング結果のクラスタ順序は何らかの意味を持っているので、概要一覧表のクラスタ名の順序もそれに従うようにした方が、ユーザが情報を探す上で都合がよい場合が多い。ユーザは一般に上から順に見て行く傾向にあるので、概要一覧表のクラスタ名の配置順序をクラスタリング結果におけるクラスタの配置順序と同じにすれば、自分の欲しい情報を探すのに便利である。

#### 【0054】

次に、概要一覧表10の或るクラスタ名をクリックしてそれにリンクされたクラスタリング結果20のクラスタ部分を表示させる際、そのクラスタを囲っている枠（クラスタ枠という）の最上部を先頭にした表示を行う。これは、もしクラスタのクラスタ名部分を基準にしてそのクラスタ名部分を先頭にした表示を行うと、表示させたときに、そのクラスタ名に対応するクラスタ要素（図4ではクラスタに含まれる文書のタイトル）が表示画面上に表示されない場合があるという不都合が生じる。たとえば、画面上にクラスタリング結果20の半導体クラスタが現れていない状態で、概要一覧表10の「半導体」というクラスタ名をクリックされたとすると、そのクリック動作とほぼ同時にクラスタリング結果20の半導体クラスタ部分の表示に移行することになるが、そのとき、そのクラスタ枠内の最上段の行に位置するクラスタ要素（この場合は、「157 半導体事業部環境方針」という文書のタイトル）が表示画面上で見えないということがしばしば生じる。

## 【0055】

これを防ぐために、クラスタ枠を基準にしてそのクラスタ枠の最上部を先頭にした表示を行うようにする。これによって、そのクラスタ枠内の最上段に位置する行のクラスタ要素が確実に表示がなされるようになる。

## 【0056】

また、さらに確実性を高めるため、それより1つ前に位置するクラスタの最終行のクラスタ要素を先頭にした表示を行うようにしてもよい。たとえば、前述の例で考えれば、概要一覧表10の「半導体」というクラスタ名がクリックされたとすると、クラスタリング結果の半導体クラスタが表示されることになるが、このとき、それよりも1つ前のデバイスクラスタの最終行のクラスタ要素（図4では、「56 デバイス-半導体-ASSP」）といったクラスタ要素を先頭にした表示を行うようにしてもよい。このように、1つ前に位置するクラスタの最終行のクラスタ要素から表示させるようにすれば、本来表示すべきクラスタのクラスタ要素は確実に表示がなされるようになる。

## 【0057】

次に、概要一覧表10に表示されるそれぞれのクラスタ名は、クラスタリング結果20におけるそれぞれのクラスタの内容によって表示の大きさや色を変える。このクラスタリング結果におけるそれぞれのクラスタの内容というのは、具体的には、それぞれのクラスタの重要度を示すもので、ユーザの与えたキーワードに対する適合度などであり、それぞれのクラスタに含まれる文書数やそれぞれのクラスタのスコアなどを用いて決める。たとえば、前述したように、それぞれのクラスタにおけるスコアの平均を求め、その平均のスコアの大きいクラスタが重要度の高いクラスタであるとすれば、重要度の高いクラスタに対する概要一覧表10のクラスタ名の表示の仕方を変える。

## 【0058】

たとえば、前述した例によれば、図4のクラスタリング結果におけるそれぞれのクラスタのなかで、半導体クラスタのクラスタスコアが最も高い値であったので、半導体クラスタに対応する概要一覧表のクラスタ名「半導体」の表示の仕方を他のクラスタのクラスタ名と異ならせる。具体的には、「半導体」というクラ



スタ名を他のクラスタ名と色を変えて表示したり、「半導体」というクラスタ名を囲っている枠だけを他よりも太くしたり、その枠内の面積を他よりも大きな面積としたり、さらにはそのクラスタ名をブリンク表示させたりすることによって、ユーザの目につきやすい表示を行う。

#### 【0059】

また、それぞれのクラスタにおいてそれぞれのクラスタに含まれる文書数の多いクラスタについても同様に、ユーザの目につきやすい表示を行うようにすることもできる。さらに、スコアの高いクラスタは色を他のクラスタとは異ならせ、文書数の多いクラスタは枠内の面積を大きくするなど、クラスタの特徴によって表示の仕方を多種多様に設定するようにしてもよい。このようにすれば、どのクラスタがどの同様な特徴を持ったクラスタであるかということが、概要一覧表10を見るだけで一目でわかる。

#### 【0060】

また、概要一覧表10においてユーザの入力したキーワードが含まれるクラスタ名は、それをユーザに知らせるために他のクラスタ名と異なった表示を行う。たとえば、図2のクラスタリング結果の例では、ユーザの与えたキーワードは「半導体」であるため、クラスタリング結果により得られたクラスタのうち、半導体クラスタはまさにキーワードそのものを含むクラスタである。

#### 【0061】

この場合、そのクラスタリング結果20により作成された概要一覧表10において、「半導体」部分は他のクラスタとは異ならせた表示として、たとえば、ブリンクさせたり、色を変えたり、両者を組み合わせるなど、ユーザの目につきやすい表示を行う。一般に、ユーザは自分の入力したキーワードと同じものがあるかを探すことが多い。したがって、概要一覧表10内に自分の与えたキーワードと同じクラスタ名が有るかどうかを一目でわかるようにすれば、ユーザにとっては自分の欲しい情報を探す上で便利なものとなる。

#### 【0062】

また、画面上にクラスタリング結果20の或る部分が表示されている状態のときに、その表示内容から或るクラスタ名をクリックすると概要一覧表10の表示

に戻るが、そのとき、概要一覧表 10 の先頭に戻って、概要一覧表 10 の先頭部分からの表示がなされるようにする方が都合がよい。

#### 【0063】

概要一覧表 10 はクラスタリング結果の概要を簡略化して表すものではあるが、場合によってはそのサイズがきわめて大きなものとなることもある。また、複数の概要一覧表が作成される場合もある。すなわち、これまでの説明では、それぞれの文書のタイトルを用い、そのタイトルによってクラスタリング処理しそれによって得られたクラスタリング結果に基づいて概要一覧表を作成する例について説明したが、クラスタリング処理は、タイトルだけでなく、URL アドレス（<http://>を取り除いた部分）などの情報を用いても行うことができる。

#### 【0064】

たとえば、図 2 のクラスタリング結果を得るために用いた多数の文書に対し、URL を用いてクラスタリング処理し、それによって得られたクラスタリング結果に基づいて概要一覧表を作成することもできる。

#### 【0065】

図 5 は図 2 と同じ文書に対し、URL アドレスによってクラスタリングされたクラスタリング結果 30 とそれに基づいて作成された概要一覧表 40 の例を示すものである。この場合、クラスタリングされることによって得られたクラスタのクラスタ名としては、クラスタリングする際の用いられたそれぞれの URL アドレス「[www.epson.co.jp](http://www.epson.co.jp)」と「[www.i-love-epson.ne.jp](http://www.i-love-epson.ne.jp)」や「その他」であり、概要一覧表 40 を構成するクラスタ名は、これらの URL アドレス「[www.epson.co.jp](http://www.epson.co.jp)」と「[www.i-love-epson.ne.jp](http://www.i-love-epson.ne.jp)」さらに「その他」が用いられる。

#### 【0066】

この例で説明したように、クラスタリング処理は、幾つもの方法によってなされる可能性があり、その場合、それぞれによって得られたクラスタリング結果に対する複数の概要一覧表が作成されることになる。

#### 【0067】

このように、複数の概要一覧表が作成される場合や、一つの概要一覧表であっても一覧表のデータ量がきわめて大きい場合、画面上でクラスタリング結果を見

ていて、それに対応する概要一覧表のクラスタ名部分に戻したとき、対応するクラスタ名部分からの表示を行うのではなく、複数の概要一覧表が存在する場合には、複数の概要一覧表のなかで先頭に位置する概要一覧表の先頭部分からの表示がなされるようにする。また、1つの概要一覧表の場合でも、その概要一覧表の先頭部分からの表示がなされるようにする。

#### 【0068】

これは、たとえば、ユーザがクラスタリング結果を見ていて、もう一度、概要一覧表に戻ってクラスタリング結果全体を見たいというような場合に対処するものである。このような場合、概要一覧表の対応するクラスタ名部分に戻ると、戻った位置が一覧表全体のどの位置かがわからなくなる可能性があるが、概要一覧表が複数存在していても、あるいは概要一覧表のデータ量が多くても、概要一覧表全体の先頭部分からの表示がなされるようにすれば、一覧表全体を見渡すことができる。

#### 【0069】

以上説明したような種々の機能を追加することによって、概要一覧表の有用性をより一層高めることができる。

#### 【0070】

以上説明したように、この実施の形態では、膨大な情報をクラスタリングし、そのクラスタリング結果の概要が一目でわかるような概要一覧表をクラスタリング結果とともに表示するようにしたので、クラスタリング結果が膨大な内容となったとしても、クラスタリング結果全体を見渡すことができ、ユーザが自分の欲しい情報を探す際の大きな助けとなる。また、その概要一覧表とクラスタリング結果との間に前述したような種々の機能を持たせることによって、ユーザの与えたキーワードに対し重要度の高いクラスタはどれであるかなどが一目でわかり、また、概要一覧表からクラスタリング結果に移行する際、あるいは、クラスタリング結果から概要一覧表に戻る際、表示する際の先頭を的確な位置に選ぶことによって、クラスタリング結果と概要一覧表とを相互に表示するような操作を繰り返し行う場合、必要な部分の表示が切れて見えない状態となったり、自分の注目している部分がどこにあるのかがわからなくなることがなくなり、能率よく快適

な操作が可能となる。

【0071】

なお、本発明は以上説明した実施の形態に限定されるものではなく、本発明の要旨を逸脱しない範囲で種々変形実施可能となるものである。たとえば、前述の実施の形態では、検索されたそれぞれの文書のタイトルを用いてクラスタリングする例について説明したが、これは、タイトルだけでなく、前述したように、URLアドレス（http://を取り除いた部分）を用いたクラスタリングも可能で、その他、更新日時（単純な時間または最近1カ月以内の更新日時）、ファイルサイズ（webページ本文のバイトサイズなど）を用いてクラスタリングすることもできる。また、これらは、単独で用いてクラスタリングするようにしてもよく、幾つかを組み合わせてもよい。このように、クラスタリング対象情報を種々選ぶことによって、それぞれに応じた特色のあるクラスタリングが行える。そして、それぞれのクラスタリング結果に基づいてそれぞれの概要一覧表を作成することができる。

【0072】

また、これまでの説明では、ある1つの汎用の検索サービスで検索された結果をクラスタリング処理する場合について説明したが、複数の検索サービスにより検索された結果をクラスタリング処理する場合にも適用できる。そして、それぞれの検索サービスによって検索された結果に基づいてクラスタリング処理し、それによって得られたクラスタリング結果に基づいて概要一覧表の作成を行うことが可能である。

【0073】

また、本発明は、クラスタリング処理を行って得られたクラスタリング結果に対し、そのクラスタリング結果によって得られたそれぞれのクラスタ間でそれぞれのクラスタの関連性を判断し、一定以上の関連性を有する少なくとも2つのクラスタを統合するクラスタマージ処理を行ったクラスタリング結果に対し、そのクラスタ結果の概要一覧表を作成することもできる。

【0074】

たとえば、図2のクラスタリング結果に対しクラスタマージ処理を行った結果

、概要クラスタと仕様クラスタがマージされたとすれば、それによって新たに「概要」と「仕様」というクラスタ名を持つクラスタが生成されることになる。この場合、そのクラスタリング結果に基づいて概要一覧表を作成すると、その新たなクラスタのクラスタ名である「概要」と「仕様」が取り出されて、その「概要」と「仕様」というクラスタ名が同じ枠内に入った概要一覧表が作成される。

#### 【0075】

また、以上説明した本発明の情報分類処理を行う処理プログラムは、フロッピーディスク、光ディスク、ハードディスクなどの記録媒体に記録しておくことができ、本発明はその記録媒体をも含むものである。また、ネットワークから処理プログラムを得るようにしてもよい。

#### 【0076】

##### 【発明の効果】

以上説明したように本発明によれば、検索された複数の文書をクラスタリング処理し、それによって得られたクラスタリング結果に対し、そのクラスタリング結果の概要が一目でわかるようなクラスタリング結果概要一覧表作成し、そのクラスタリング結果概要一覧表をクラスタリング結果とともに表示するようにしたので、クラスタリング結果が膨大な内容となったとしても、クラスタリング結果全体のおおよその内容を把握することができ、ユーザが自分の欲しい情報を探す際の大きな助けとなる。また、クラスタリング結果全体のおおよその内容を把握できることから、ユーザは自分の欲しい情報を効率よく探すことができるばかりでなく、思わぬ情報の存在を発見することもでき、新規情報の発掘も容易に行えるようになる。

#### 【0077】

また、クラスタリング結果とクラスタリング結果概要一覧表とを相互にリンクすることで、クラスタリング結果概要一覧表からクラスタリング結果の対応するクラスタ部分に簡単に移行することができ、また、その状態から、クラスタリング結果概要一覧表にも容易に戻ることもできるので、クラスタリング結果が膨大な内容であっても、このような操作を何回か行うことで、やがては欲しい情報の入っているようなクラスタを探すことができ、自分の欲しい情報を効率よく探すこと

ができる。

【0078】

また、クラスタリング結果概要一覧表の表示を行う際、前記クラスタリング結果のそれぞれ対応するクラスタの重要度に応じて、クラスタリング結果概要一覧表内のクラスタ名の表示の仕方を変えるようにしたり、ユーザの与えたキーワードを含むクラスタ名の表示の仕方を他のクラスタ名とは異ならせるようにすることで、自分の欲しい情報がどこに入っているかがそのクラスタリング結果概要一覧を見るだけでおおよその見当を付けることができ、自分の欲しい情報を効率よく探すことができる。

【図面の簡単な説明】

【図1】

本発明の情報分類装置の実施の形態を説明する構成図である。

【図2】

ある検索サービスで検索された複数の文書をクラスタリングした結果の一例を示す図である。

【図3】

本発明の情報分類処理手順を概略的に説明するフローチャートである。

【図4】

図2で示されたクラスタリング結果とその概要一覧表を表示した例を示す図である。

【図5】

URLアドレスでクラスタリング処理して得られたクラスタリング結果とその概要一覧表を表示した例を示す図である。

【符号の説明】

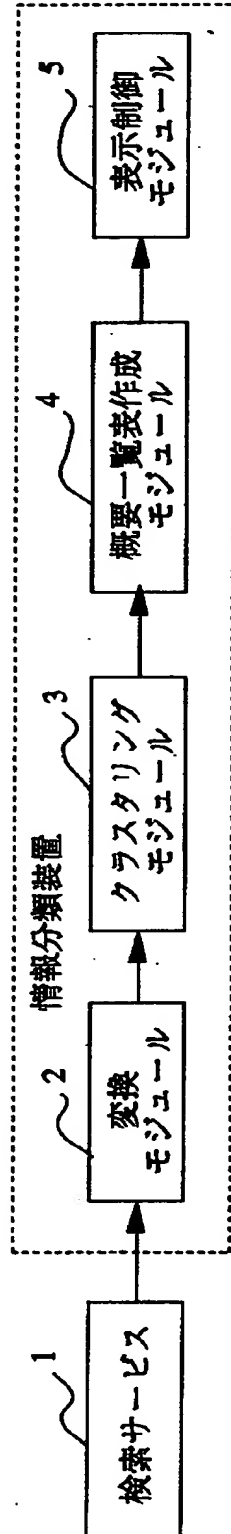
- 1 検索サービス
- 2 変換モジュール
- 3 クラスタリングモジュール
- 4 概要一覧表作成モジュール
- 5 表示制御モジュール

1 0 概要一覧表

2 0 クラスタリング結果

【書類名】 図面

【図 1】

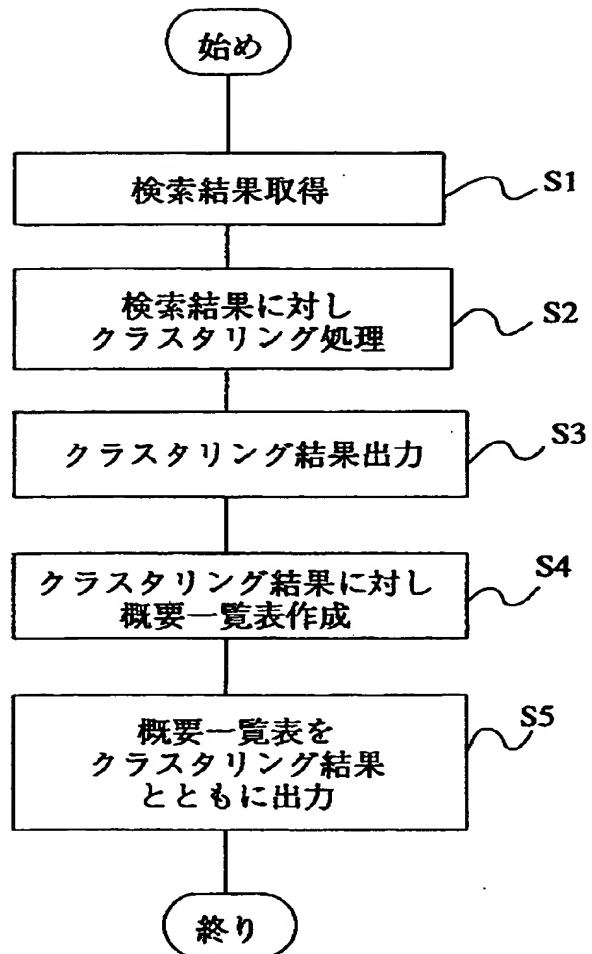




【図 2】

クラス名	文書数	スコア	文書のタイトル
概要	16		133 東北エプソンの概要 49 LP-7000 仕様概要 38 LP-8000 仕様概要 38 LP-8400 仕様概要 34 LP-700 仕様概要 30 セイコーエプソン会社概要 30 LP-9200 仕様概要 28 LP-9200PS2 仕様概要 28 LP-4300PS 仕様概要 27 LP-4300 仕様概要 26 LP-8600 仕様概要 24 LP-8200 仕様概要 23 LP-710 仕様概要 23 LP-800 仕様概要 23 LP-900 仕様概要 22 LP-500 仕様概要
LP	16		49 LP-7000 仕様概要 38 LP-8000 仕様概要 38 LP-8400 仕様概要 34 LP-700 仕様概要 30 LP-700S Spec Sheet & Option 30 LP-9200 仕様概要 28 LP-9200PS2 仕様概要 28 LP-4300PS 仕様概要 27 LP-4300 仕様概要 26 LP-8600 仕様概要 24 LP-8200 仕様概要 23 LP-710 仕様概要 23 LP-800 仕様概要 23 LP-900 仕様概要 22 LP-500 仕様概要 7 エスパーレーザー LP-4300
仕様	14		49 LP-7000 仕様概要 38 LP-8000 仕様概要 38 LP-8400 仕様概要 34 LP-700 仕様概要 30 LP-9200 仕様概要 28 LP-9200PS2 仕様概要 28 LP-4300PS 仕様概要 27 LP-4300 仕様概要 26 LP-8600 仕様概要 24 LP-8200 仕様概要 23 LP-710 仕様概要 23 LP-800 仕様概要 23 LP-900 仕様概要 22 LP-500 仕様概要
デバイス	9		117 デバイス 半導体 CARD-PC 115 デバイス 半導体 メモリ 111 電子デバイス ASIC 111 デバイス 半導体 PCカード製品 101 デバイス コラボレーション 90 電子デバイス 半導体 マイコン 82 EPSON 電子デバイス新製品 66 EPSON 電子デバイスお問い合わせ 56 デバイス 半導体 ASSP
半導体	7		157 半導体事業部環境方針 117 デバイス 半導体 CARD-PC 115 デバイス 半導体 メモリ 111 電子デバイス ASIC 111 デバイス 半導体 PCカード製品 90 電子デバイス 半導体 マイコン 56 デバイス 半導体 ASSP
電子	4		111 電子デバイス ASIC 90 電子デバイス 半導体 マイコン 82 EPSON 電子デバイス新製品 66 EPSON 電子デバイスお問い合わせ

【図 3】



【図 4】

クラスタリング結果概要一覧					
概要	LP	仕様	デバイス	半導体	電子
スコア 文書のタイトル 検索結果					
概要	16	133 東北エプソンの概要 49 LP-7000 仕様概要 38 LP-8000 仕様概要 38 LP-8400 仕様概要 34 LP-700 仕様概要 33 セイコーエプソン会社概要 30 LP-9200 仕様概要 28 LP-9200PS2 仕様概要 28 LP-8300PS 仕様概要 27 LP-8300 仕様概要 26 LP-8600 仕様概要 24 LP-8200 仕様概要 23 LP-710 仕様概要 23 LP-800 仕様概要 23 LP-900 仕様概要 22 LP-500 仕様概要			
LP	16	49 LP-7000 仕様概要 38 LP-8000 仕様概要 38 LP-8400 仕様概要 34 LP-700 仕様概要 30 LP-700S Spec Sheet & Option 30 LP-9200 仕様概要 28 LP-9200PS2 仕様概要 28 LP-8300PS 仕様概要 27 LP-8300 仕様概要 26 LP-8600 仕様概要 24 LP-8200 仕様概要 23 LP-710 仕様概要 23 LP-800 仕様概要 23 LP-900 仕様概要 22 LP-500 仕様概要 7 エスパーレーザ LP-8300			
仕様	14	49 LP-7000 仕様概要 38 LP-8000 仕様概要 38 LP-8400 仕様概要 34 LP-700 仕様概要 30 LP-9200 仕様概要 28 LP-9200PS2 仕様概要 28 LP-8300PS 仕様概要 27 LP-8300 仕様概要 26 LP-8600 仕様概要 24 LP-8200 仕様概要 23 LP-710 仕様概要 23 LP-800 仕様概要 23 LP-900 仕様概要 22 LP-500 仕様概要			
デバイス	9	117 デバイス 半導体 CARD-PC 115 デバイス 半導体 メモリ 111 電子デバイス ASIC 111 デバイス 半導体 PCカード製品 101 デバイス コラボレーション 90 電子デバイス 半導体 マイコン 82 EPSON 電子デバイス新製品 66 EPSON 電子デバイスお問い合わせ 56 デバイス 半導体 ASSP			
半導体	7	157 半導体事業部環境方針 117 デバイス 半導体 CARD-PC 115 デバイス 半導体 メモリ 111 電子デバイス ASIC 111 デバイス 半導体 PCカード製品 90 電子デバイス 半導体 マイコン 56 デバイス 半導体 ASSP			
電子	4	111 電子デバイス ASIC 90 電子デバイス 半導体 マイコン 82 EPSON 電子デバイス新製品 66 EPSON 電子デバイスお問い合わせ			

【図 5】

クラスタリング結果概要一覧		
www.epson.co.jp	www.i-love-epson.co.jp	その他

クラスタ名	文書数	検索結果
www.epson.co.jp	31	<p>半導体事業部 環境方針 先輩からのメッセージ-テキスト http://www.epson.co.jp/epson/ichisaiyou/taooc/ http://www.epson.co.jp/epson/ichisaiyou/taoc/ 東北エプソンの概要 「vividly NOTE」 http://www.epson.co.jp/epson/ichisaiyou/taoc/ 東北エプソン 先輩からのメッセージ Lamtec社とセイコーエプソン、半導体ビジネス 東北エプソン 企業データ デバイス-半導体-CARD-PC 半導体環境方針 デバイス-半導体-メモリ http://www.epson.co.jp/epson/ichisaiyou/taoc/ http://www.epson.co.jp/epson/ichisaiyou/taoc/ 製品内容 電子デバイス-半導体-ASIC デバイス-半導体-PCカード製品 http://www.epson.co.jp/epson/japanese/ichisaiyou/ 東北エプソン 基本ポリシー 西田5号機竣工 デバイス コラボレーション ページ EPSON ELECTRONIC DEVICES INDEX 東北エプソン株式会社 環境方針 ムッシュ 上海に新会社Shanghai Epson Electronicsを設 EPSON Ecology 電子デバイス-半導体-マイクロコンピュー SEIKO EPSON GROUP http://www.epson.co.jp/epson/japanese/ichisaiyou/ EPSON 電子デバイス新製品</p>
www.i-love-epson.co.jp	14	<p>LP-7000 仕様概要 LP-8000 仕様概要 LP-8400 仕様概要 LP-700 仕様概要 LP-9200 仕様概要 LP-9200P32 仕様概要 LP-830FS 仕様概要 LP-8300 仕様概要 LP-8500 仕様概要 LP-8200 仕様概要 LP-710 仕様概要 LP-800 仕様概要 LP-900 仕様概要 LP-500 仕様概要</p>
その他		<p>自然環境問題：オゾン層破壊 オゾン層とは RX-Server：FP-430</p>

【書類名】 要約書

【要約】

【課題】 汎用の検索サービスによって検索された多数の文書をクラスタリング処理した場合であっても、それによって得られたクラスタリング結果内容が膨大な量となることもあり、クラスタリング結果全体を把握してその中から自分の欲しい情報を探し出すのは多くの労力を必要とする。

【解決手段】 汎用の検索サービス 1 で検索された複数の検索結果をクラスタリングモジュール 3 が取得して、検索結果をクラスタリング処理し、さらに、そのクラスタリング結果に対し、クラスタリング結果概要一覧表作成モジュール 4 がそれぞれのクラスタのクラスタ名による概要一覧表を作成する。そして、表示制御モジュール 5 によって、その概要一覧表をクラスタリング結果とともに表示させる。

【選択図】 図 1

出 願 人 履 歴 情 報

識別番号

[ 0 0 0 0 0 2 3 6 9 ]

1. 変更年月日

1 9 9 0 年 8 月 2 0 日

[ 変更理由 ]

新規登録

住 所

東京都新宿区西新宿 2 丁目 4 番 1 号

氏 名

セイコーエプソン株式会社